

5

**TECHNIQUE TO IDENTIFY INTERESTING PRINT ARTICLES FOR
LATER RETRIEVAL AND USE OF THE ELECTRONIC VERSION OF THE
ARTICLES**

10

Field of the Invention

The present invention relates generally to information organization and retrieval, and more particularly to facilitating organization and retrieval of electronic information.

15

Background of the Invention

09738209.121500

The amount of information generated and available in today's environment tends to be overwhelming. The most important information to an individual is the information that he/she has examined and determined to be useful to him/her. Today, most important information is available or passes through electronic form, even though it may ultimately be distributed primarily in printed form. When information is received electronically, such information takes little physical space and may be indexed manually or using an automatic index and classification system so that it may be stored and retrieved at a later time. However, when the information is printed, it is often inconvenient or difficult to connect the printed information with its electronic counterpart. This may be true even if the individual printed the information from its electronic form, such with a personal printer.

Often, the individual receiving the information may wish to read it and keep it for future reference, but does not have the opportunity to read the information when it is received. For example, people may quickly review a magazine and bend down page corners of interesting articles or rip out the articles, often reading only headlines or a one to two paragraphs of the article. Items of interest, such as travel, ornithology, enology, viticulture and gardening may be placed on a countertop in a "holding"

pattern, stacked in piles, put in boxes and folders, and moved several times before either being filed in a huge, often unsorted, pile or folder to be read at some future date, or filed in the trash.

Many people have no time for categorizing and organizing this saved
5 information, and it ends up in a growing number of piles, often with years worth of unsorted information that is difficult to access in an efficient manner. Hence, the original purpose of saving the article for later reference or reading is defeated. For example, consider the subject of travel. An individual may sort all travel articles into a folder; if not, the travel articles are all mixed up in multiple places. Among the
10 mass of saved articles, there may be an article on "The Best Places to Stay in Hawaii", another article on "Maui's Best Molokini Excursions", and many other articles on Europe, Tahiti, and the Virgin Islands.

If the family desires to go to Maui, finding the needed information in the saved articles may present a large task. First, the general location of the travel articles
15 in piles or folders must be determined. Then, if the articles are categorized, the individual must sort through the articles. If the articles were classified rather definitively, under what title would the articles on the places to stay in Hawaii be stored? In a Hawaii folder? Are there sub-folders for each island? If the articles are not categorized, it will be even more difficult since the individual will have to sort the
20 entire pile or box/boxes of articles. Clearly, organizing and retrieving the desired information that may be useful in the future is not a trivial task and is time-consuming.

Thus, there is a need for a method and system to facilitate organizing and retrieving both printed and electronic information efficiently.

25

Summary of the Invention

The present invention provides a method for efficient information storage and retrieval of information. The method includes the step of scanning/selecting/capturing a selected portion of text of the information wherein the
30 selected portion of text captured is a fragment of information content that may be used as an approximate identifier or "key" for the entire information and where the information may be stored and accessed electronically. Then, the selected portion of

the text or "key" is placed in an electronically available index/directory to facilitate retrieval of the information from a storage medium having the information stored electronically.

- The present invention may also be utilized as a system for efficient
- 5 information storage and retrieval of information. In this embodiment, the invention includes an index/directory information retrieval unit, an index/directory storage unit, and a storage medium. The index/directory information retrieval unit is arranged to send a portion of text of the information to an index/directory storage unit. The index/directory information retrieval unit is used for scanning/selecting/capturing the
- 10 portion of text of the information wherein the portion of text is a fragment of information content that may be used as an approximate identifier or "key" for the entire information. The information is stored and accessed electronically. The index/directory storage unit is arranged to receive and store the key from the index/directory information retrieval unit. The index/directory storage unit is used for
- 15 storage of the key in an electronically available index/directory to facilitate retrieval of the information from a storage medium that is accessible electronically. The index/directory is arranged to be searched to identify information in the storage medium that corresponds to the key. Where the information is already in electronic form, the information is stored in a storage medium and made available for search and
- 20 retrieval based on text fragments. Where the information is printed, the information is scanned and processed to facilitate search and retrieval based on text fragments and also ideally retain the form and presentation of the information to provide an electronic copy of the information which is then stored in a storage medium and made available for search and retrieval based on text fragments. For example, the
- 25 information may be stored in a personal or organization-wide database, or may be from a commercially accessible database. Typically, a computer/user unit is coupled to or arranged to receive/send information from/to the index/directory unit and the storage medium and is used to request retrieval of information based on the key.

- In one embodiment, the invention may be implemented by a computer-
- 30 readable medium that has computer-executable instructions for performing steps for efficient information storage and retrieval of information. The computer-executable instructions perform the steps described above.

09738209-121500

Brief Description of the Drawings

Figure 1 is a schematic showing one embodiment of a system in accordance with the present invention.

5 Figure 2 is a flow chart of steps of a method for efficient information storage and retrieval of information in accordance with the present invention.

Figure 3 is a block diagram of a system for efficient information storage and retrieval of information in accordance with the present invention.

10 Figure 4 is a schematic of an embodiment of the present invention in which scanned keys are stored in an index/directory and the keys are associated with multiple information repositories containing electronic versions of articles or information.

Detailed Description of the Preferred Embodiments

15 The present invention provides a method and device for organizing and retrieving both printed and electronic information in which the information is identified by the person viewing the information by selection of a portion of the text, using the selection as a close-to-unique identifying "key", and where the information is not in electronic form, obtaining an electronic copy of the information and storing
20 the information electronically, thus allowing one to locate the information quickly using the key. The invention applies to any given package of information from small to large that includes written or spoken language that can be processed as text sequences, and may also include other forms of information such as charts, pictures, audio, and video. For example, common information packages include an article and
25 its associated charts or graphics, a book, a document, a webpage, a letter, a resume, a form, a press release, and so forth. While the term "article" is used for clarity of writing, those skilled in the art will understand it to be representative of many kinds of information packages.

For example, as shown in Figure 1, for a printed article that has significance to
30 a user, the user scans/selects/captures a small amount of identifying text 102, e.g., one or two sentences from the article. This small amount of text becomes a close-to-unique identifier for the article, and may be used as the key value entered into systems

09738209.121500

that index, classify, associate, and retrieve information. No special identifier, such as a bar code or article number, is needed. Because of the richness of human language, the one or two sentences are close to a unique key for the article.

A scanner, such as are available commercially presently, may be used to
5 implement the capturing of text. The scanner captures the text 102, either as an image that facilitates later Optical Character Recognition (OCR) on a different system or doing OCR directly to produce the text. The information captured by the scanner is sent to one or more systems that process the scan. The scanner may be coupled to the system in any desired known manner such as, for example, using infrared (IR)
10 transmission, a synchronizing system, a Universal Serial Bus (USB) connection or wireless or the like. While any scanner may be used, the use of small hand scanners that do OCR onboard the scanner, such as the C-Pen 600, which is available from C Technologies AB, Ideon Research Park, Scheelevägen 15, SE-223 70 LUND, Sweden, is quite convenient. The C-Pen 600 is as small as a highlighter pen and is
15 easy to use on any size and shape paper, making scanning especially practical and easy.

Given the richness of human language, the small amount of text typically is a close-to-unique identifier of the complete article from which it was excerpted and can be used as a key. The key is stored in an index or directory 104. In one embodiment of
20 the invention, when a new key is added to the index/directory and found to be in use already, the system sends a message to the user that the key is already in use. Generally, this will not be a problem, as the number of matching articles will be small, but it may train the user to scan or elect longer text fragments in the future. The key may be stored in an index/directory 104, and the index/directory may be used
25 to locate the article or set of articles that match the key. Where desired, the index/directory containing the keys may be managed by a personal database, organization-wide database, or service operated by other organizations.

Where the article is in electronic form, the article may be stored in an electronic database. In one embodiment of the present invention, the article in toto is
30 stored electronically in a database 106 to improve responsiveness and reduce dependence for article access on remotely operated systems. Clearly, this may be accomplished by scanning the article, or may be accomplished by utilizing

09738209.121500

information electronically available from commercial companies such as LexisNexis®, NorthernLight®, and the like, which have undertaken to collect, index, classify, and make available electronic versions of thousands of print collections, and information printed from electronic content located on the Internet.

Both the keys and the articles may be used to provide useful classifications and indices and to allow search, discovery and retrieval of information. Systems may be used to provide personalized keyword indices, subject and other classifications and to allow for searching through a set or subset of the information identified as being of interest. Optionally, systems may recommend relevant and/or related to the topic matters of the information being captured or retrieved such as URLs, maps, books, and other related articles. Systems may also be used to provide desired information such as personal notes for the user on a particular topic or pictures.

In practice, as information is read, users no longer need to rip out articles or fold corners of articles of interest and stack them in piles or organize them in folders.

Instead, users may scan a few lines of text from the articles, for example, using a small hand scanner, where the scanned lines are used as keys and occasionally upload the scanned keys to a system that may locate and process the electronic version of the articles referenced by the keys. Thus, at a future time, when the person is interested in retrieving the articles, key words may be entered, for example on a personal computer 108, and the search conducted across the set of articles 110 that have been identified by the user as being of interest. In the example shown in Figure 1, the clipping list of articles identified by the user as being of interest contains three articles: 1. Excursions to Molokini from Maui; 2. The best places to stay in Hawaii; and 3. Maui's Farmer's Markets. From the search results list, the more recent articles or information that are related or similar, may be noted, and the desired material may be selected. The results may be delivered to a screen or whatever output device 116 is selected. For example, where desired, the articles may be sent to a printer 112 to be printed out 114 for reference by the user.

The present invention provides for, upon a user's reading an article in print or electronically, wherein the article is of interest to the user, the user's scanning/selecting/capturing a small selected amount of text. As used hereinafter, and for purposes of claim construction, the term "scanning/selecting/capturing" refers to

substantially equivalent processes by which a user who wishes to retrieve at a later date, an article or document of interest, first identifies or selects some small portion or aspect of the article or document to be used a key, or index, to retrieve the entire document, which is electronically stored elsewhere. In the case of documents or printed articles, the selected key can be "scanned" and turned into computer-storable data using an optical scanner. In the case of a document in an electronic form, the portion of the electronic document selected to be a key can be simply "selected" much as text is selected using a word processor by holding down certain keys and dragging a mouse pointer icon over portions of text on screen. In the case of graphics e.g. a portion of a picture might be "captured" or a spoken description thereof might be "captured." For purposes of claim construction, any process or data that provides for the digitization of information from or describing a document, and that is used as an index to the stored versions of such documents, is considered to be equivalent to scanning or selecting or otherwise capturing ("scanning/selecting/capturing") a portion of the document as a key. The scanned text is used as a close-to-unique identifier or key for the article. When the key is uploaded to a processing system that has access to a collection of electronic versions of printed information as well as electronic information, the key may be used to locate the electronic version of the printed/electronic article. Once the electronic version of the article is available, indices and classifications of the article may be created, manually or automatically, that allow for rapid search and retrieval using words, concepts, classifications, and the like that apply to the whole article and not just the key. Where desired, such searches may be limited only to articles that the user has seen and identified to be of interest, and hence has a known level of interest to the person. However, the present invention may also include, upon searching, providing a list of associations with other articles that may be relevant to the user, but which the user has not yet reviewed yet may be likely to desire to access.

As shown in Figure 2, the present invention provides a method for efficient information storage and retrieval of information. The method includes the steps of: scanning/selecting/capturing 202 a selected portion of text of the information wherein the selected portion of text scanned typically is a close-to-unique identifier of the text, and serves as a "key" when the information is stored and accessed electronically; and

placing 204 the key in an electronically available index/directory to facilitate retrieval of the information. A close-to-unique identifier is defined as a selected portion of text that the user chooses as highly likely, when used to identify information to be retrieved, to retrieve the text associated therewith. In practice, for information across

5 a broad range of academic, business, and personal content, suitable text can be very easily chosen, such as the first sentence or first line of text of an article. Typically, the information is in electronic form and the method may further include a step of storing

206 the electronic information in a storage medium and making the electronic information available for search and retrieval using the key. - Where the information

10 is printed, the method 200 may include the steps of scanning 208 the information to provide an electronic copy of the information and storing 210 the electronic copy of the information in a storage medium and making the electronic copy available for search and retrieval using the key. Steps 206, 208, and 210 are typically done by someone other than the user. The information may be stored in a personal or

15 organization-wide database, or may be obtained from a commercially accessible database. In addition, the search mechanism may be arranged to search hard drives or any selected storage medium. The method may further include, upon the user requesting retrieval of information associated with the key, providing 212 the information associated therewith. Where desired, further related/relevant information

20 may also be provided 214 that is correlated with the key, with associated information, or both. In addition, in a forward mode of use, where a time period exists between a printed publication and actual storage of an electronic version of the information in an available database, the user may clip and hold 216 a key so that the information may be accessed at a later date. Thus, the selected portion of text serves as a "clip and

25 hold" key for accessing information that is electronically available at a later date. Where desired, the method may be implemented by a computer-readable medium 218 with computer-executable instructions to carry out the steps of the method. In one embodiment, the computer-executable instructions may be used to provide an index/directory for incoming information such as business forms or resumes, and the

30 keys in the index/directory are placed on a web page or database that is accessible from distributed locations so that documents associated with the keys may be accessed from many locations. In such a system, the documents may be stored in

09738209.121500

distributed locations. Where desired, clearly a user may simply upload keys for his own account and store such keys on an intermediate index/directory for his own use.

As shown in Figure 3, the present invention may include a system for efficient information storage and retrieval of information. Typically, the system includes an index/directory information retrieval unit 302 (typically a scanner), an index/directory storage unit 304 and a storage medium 306. The index/directory information retrieval unit 302 is arranged to be utilized for scanning, selecting or capturing a portion of text of the information and sending the portion of text to a index/directory storage unit 304. The portion of text typically is a close-to-unique identifier of the complete information from which it was excerpted and serves as a "key". The information is stored and accessed electronically. The index or directory storage unit 304 is arranged to receive and store the key from the index or directory information retrieval unit 302 and may also test the referenced information itself. The key is placed in an electronically available index/directory in the index/directory storage unit where the keys may be used to facilitate retrieval of the information from which the key was selected from the storage medium 306 in which the information is accessible electronically. The storage medium 306 is used to store information that is accessible electronically, and the index/directory storage unit 304 is arranged to be identify information in the storage medium 306 that corresponds to the key. The information is typically in electronic form. However, when the information is in printed form, the information is typically scanned and processed and stored in electronically accessible form in the storage medium 306. The information may be stored in a personal or organization-wide database, or the information may be obtained from a commercially accessible database. The steps may further include, upon the user requesting retrieval of information associated with the key, providing the information associated therewith. Where desired, further related/relevant information may also be provided that is correlated with the key. Also, for use in a forward mode, where a time period exists between a publication and actual availability of an electronic version of the information in an archive, the user may clip and store a key so that the information may be accessed at a later date. Hence, the selected portion of text is utilized as a "clip and hold" key for accessing information that is electronically available at a later date. Either a separate database may be used for such clip and hold keys, or the key

09738209.121500

may include an indicator that shows that the information is to be accessed at a later date. Typically, a computer/user unit 310 is coupled to or arranged to receive/send information from/to the index/directory unit 304 and the storage medium 306 and is used to request retrieval of information based on the key. The user may request, for example, that the information be sent to a printer 308 for generating a printed copy of the information.

The keys and/or information to be stored electronically may be transmitted using an infrared connection, a wired connection, and/or a telephone connection. The keys and/or information may be sent to a personal computer or other device having a storage medium. A camera using a wireless, wired, or infrared connection may be arranged to provide information to the system.

Figure 4 is a schematic of an embodiment of the present invention in which scanned keys are stored in an index/directory and the keys are associated with multiple information repositories containing electronic versions of articles or information. Typically, a hand scanner 404 may be used to scan the "key" (scanned text 402). The key is then added to an index/directory of keys 406, which may be generated over a period of time. The index/directory of keys 406 may be stored locally or in a central location or distributed across locations but integrated using computer-executed instructions. The keys may be used to locate articles of interest or information, typically automatically fetching the articles and indexing or categorizing them. The index/directory of keys 406 may be associated with one information repository or multiple information repositories that store an electronic version of the article or information. The information is typically indexed for fast searching and classified for category browsing, improved search, and other uses. Where desired, the system may retrieve the articles that match the key or keys and build a local or personal index of all the articles so that the articles may be searched rapidly by a user using search words. Thus, the user may utilize the system to identify articles or browse a directory of articles that is generated without using the keys.

Although the present invention has been described in relation to particular preferred embodiments thereof, many variations, equivalents, modifications and other uses will become apparent to those skilled in the art. It is preferred, therefore, that the

05738209-121500